Structure of translation initiation factor 5A from *Pyrobaculum aerophilum* at 1.75 Å resolution

Thomas S Peat[1]*, Janet Newman[1], Geoff S Waldo[1], Joel Berendzen[2] and Thomas C Terwilliger[1]*

Background: Translation initiation factor 5A (IF-5A) is reported to be involved in the first step of peptide bond formation in translation, to be involved in cell-cycle regulation and to be a cofactor for the Rev and Rex transactivator proteins of human immunodeficiency virus-1 and T-cell leukemia virus I, respectively. IF-5A contains an unusual amino acid, hypusine (N-epsilon-(4-aminobutyl-2-hydroxy)lysine), that is required for its function. The first step in the post-translational modification of lysine to hypusine is catalyzed by the enzyme deoxyhypusine synthase, the structure of which has been published recently.

Results: IF-5A from the archebacterium *Pyrobaculum aerophilum* has been heterologously expressed in *Escherichia coli* with selenomethionine substitution. The crystal structure of IF-5A has been determined by multiwavelength anomalous diffraction and refined to 1.75 Å. Unmodified *P. aerophilum* IF-5A is found to be a β structure with two domains and three separate hydrophobic cores.

Conclusions: The lysine (Lys42) that is post-translationally modified by deoxyhypusine synthase is found at one end of the IF-5A molecule in a turn between β strands β4 and β5; this lysine residue is freely solvent accessible. The C-terminal domain is found to be homologous to the cold-shock protein CspA of *E. coli*, which has a well characterized RNA-binding fold, suggesting that IF-5A is involved in RNA binding.

Addresses: [1]Life Sciences Division, Los Alamos National Laboratory, Mail Stop M888 and [2]P-21, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA.

*Corresponding authors.
E-mail: peat@proi1.lanl.gov
terwilliger@lanl.gov

Key words: IF-5A, MAD, proteomics

Received: **20 July 1998**
Revisions requested: **5 August 1998**
Revisions received: **13 August 1998**
Accepted: **14 August 1998**

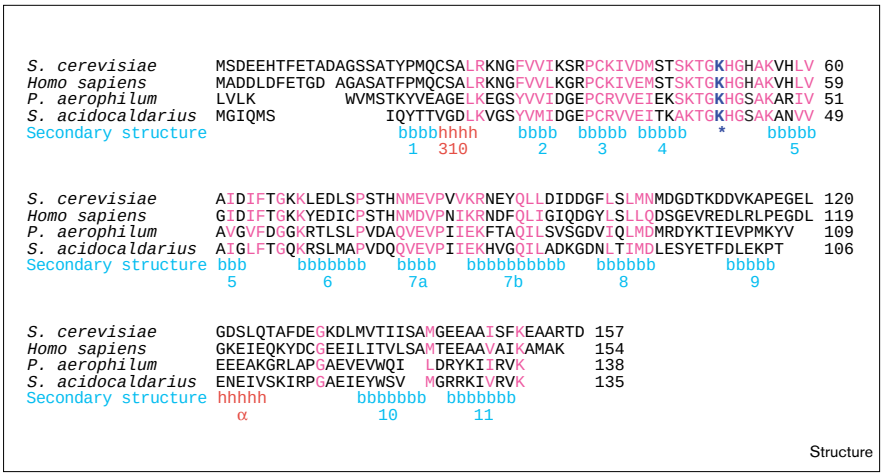
Structure 15 September 1998, **6**:1207–1214
http://biomednet.com/elecref/0969212600601207

Introduction

Initiation factor 5A (IF-5A) is an essential protein in eukaryotic cells and the modification of a specific lysine residue to hypusine is required for its function and for cell viability [1]. IF-5A is the only known protein to be modified in this way and this is one of the most specific post-translational modifications to occur in eukaryotic and archebacterial cells. The sequence surrounding the hypusine residue (Hyp) is highly conserved among eukaryotes and the 11 residues surrounding the Hyp are invariable: Ser-Thr-Ser-Lys-Thr-Gly-Hyp-His-Gly-His-Ala-Lys. The eukaryotic IF-5A (eIF-5A) homologs are not only similar at the sequence level, but can substitute for each other functionally as well [2]. It has been shown that the IF-5A proteins from the slime mold *Dictyostelium discoideum*, from alfalfa (*Medicavo sativa*) and from humans can all be hypusinated and substitute for the yeast (*Saccharomyces cerevisiae*) TIF51A *in vivo*. It was found that archebacterial IF-5A (aIF-5A, Carl Woese notation [3]) from *Sulfolobus acidocaldarius* did not substitute for the yeast gene *in vivo* and was not hypusinated. *S. acidocaldarius* aIF-5A has 8 of the 12 invariant residues in the region around the modified site (Figure 1, sequence alignment). As aIF-5A was produced in yeast cells but not modified, this lack of modification at the Lys residue is probably due to non-recognition of the archebacterial peptide substrate by the yeast deoxyhypusine sythase [2].

Eukaryotic IF-5A is a ubiquitous protein found in the cytoplasm and a large percentage is found associated with endoplasmic reticulum membranes [4]. The protein was originally isolated from a high-salt wash of rabbit reticulocyte ribosomes and was thought to be involved in translation initiation [5]. In an *in vitro* assay for translation initiation, eIF-5A can enhance the synthesis of methionyl-puromycin, suggesting that this protein is involved in the formation of the first peptide bond in translation [6]. Other experiments have also suggested that eIF-5A is involved in translation initiation and in cell-cycle regulation [7]. Experiments in yeast have shown that although deletion of both the eIF-5A genes (*TIF51A*, *TIF51B*) is lethal, depletion of eIF-5A in the cell slows the first round of protein synthesis by about 30% [8]. A temperature-sensitive mutation in the *TIF51A* gene also showed only a modest decrease in peptide synthesis in the cell when shifted to the non-permissive temperature [9]. This suggests that eIF-5A may not be directly involved in peptide synthesis but may be involved in another aspect of cell metabolism that has an indirect effect on translation.

Figure 1



A CLUSTAL W sequence alignment of four IF-5A homologs: yeast (*Saccharomyces cerevisiae*), human (*Homo sapiens*), *Pyrobaculum aerophilum*, and a second archebacterium, *Sulfolobus acidocaldarius* [34]. Residues that are highly conserved between these proteins are highlighted in magenta and the strictly conserved lysine (residue 42 in the *P. aerophilum* sequence and residue 50 in the human sequence) that is hypusinated is colored in blue and denoted with a star beneath the sequence. The alignment has 23 strictly conserved residues between the four sequences presented, with the vast majority of these being in the N-terminal domain. The secondary structure of the protein is denoted below the aligned sequences, with the cyan-colored 'b's referring to β strands and the red 'h's referring to helices (either 3_{10} or α helices).

Zuk and Jacobson [9] suggest that the role of eIF-5A in cell metabolism may be in mRNA turnover, as mutations in *TIF51A* decreased the turnover rates of several well characterized mRNAs in yeast. mRNA turnover naturally has a role in translation and is a precise process involving specific *trans*-acting factors and *cis*-acting sequences. Translation also affects the rate of mRNA turnover in many ways: several factors associated with the ribosome affect the decay process; premature translation termination can enhance mRNA decay and inhibition of translation can slow mRNA decay; 3′-untranslated region instability elements can effect mRNA translational activity; and metabolism of the poly(A) tail is a rate-limiting step in the decay of some mRNAs and yet this structure is involved in translation initiation [10]. Zuk and Jacobson's data suggest that eIF-5A may actually be involved in the stabilization of mRNAs downstream of the decapping process but upstream of the 5′–3′ exonuclease process.

Inhibition of hypusine formation in eIF-5A inhibits the start of DNA replication and entry into S phase, which is crucial for cell proliferation. This observation has led to the suggestion that eIF-5A is involved in the translation of specific mRNAs that are responsible for cell-cycle regulation. At least two translationally controlled enzymes that are crucial for cell proliferation are found to be dependent on the post-translational modification of eIF-5A by deoxyhypusine synthase [11]. eIF-5A is also found to be a cofactor for the functioning of two viral transactivator proteins, Rex and Rev, which are required for the expression of the structural proteins of their respective viruses, T-cell leukemia virus and human immunodeficiency virus [12]. The suggestion here is that the interaction of eIF-5A with Rex/Rev promotes the association of specific viral mRNAs with ribosomes, increasing the stability or the translation rate of these mRNAs.

The structure of the archebacterial IF-5A (aIF-5A) from *Pyrobaculum aerophilum* was determined as part of a pilot proteomics project currently under way in the Los Alamos and UCLA structural biology laboratories. The open reading frame (ORF) for this protein was cloned out with approximately 40 other ORFs from *P. aerophilum* genomic DNA. These ORFs were tested for heterologous protein expression in *E. coli* and those proteins that expressed well were purified and tested for crystallization. These initial trials yielded well-diffracting crystals of the aIF-5A.

Results

The structure of a selenomethionyl derivative of aIF-5A was determined using multiwavelength anomalous diffraction (MAD) from a three wavelength data set collected at beamline x8c, Brookhaven National Laboratory. There are three well-ordered selenomethionines (Met94, 96, and 106) in the protein structure and these seleniums were used to phase the electron density maps. The seleniums of residues 94 and 96 are ~4.8 Å apart and all three of these methionine residues are somewhat solvent exposed in the C-terminal domain. A fourth methionine near the N terminus (Met7) is not as well ordered and was not used in the phasing of the maps. The model is missing density for the first three amino acids, and the density is weak for residues 4–9. The density is also somewhat weak in the turn between strands β4 and β5, residues 40–44, including the normally post-translationally modified Lys42. Although the protein was expressed with eight extra amino acids at the C terminus (Gly-Ser-His_6), no density is seen after the added Gly139 residue. Therefore there are 136 total residues in the model, with eight modeled in alternative conformations, and little sidechain density for seven residues. This model also includes 143 water molecules. The final R_{free} value is 23.6%, R_{cryst} = 21.4%, with 5% reserved for the test set, no sigma cutoff and good stereochemistry (Table 1).

Table 1

Data statistics.

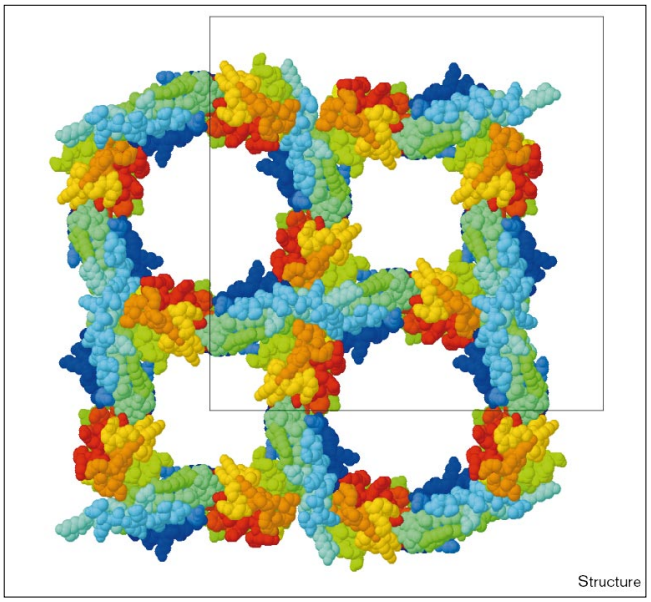
Data collection statistics* 3λ experiment, 50.0–1.75 Å	Unique reflections	Completeness (%)†	R_{sym} (%)‡	Redundancy
λ1 0.9790 (edge)	21,680	99.9 (99.1)	6.5 (26.2)	5.8
λ2 0.9788 (peak)	21,537	99.8 (99.3)	5.3 (23.0)	7.9
λ3 0.9600 (remote)	20,648	96.3 (76.0)	4.3 (24.2)	5.1
Refinement statistics§				
Resolution range	R values (R_{cryst}/R_{free})#	<B value>	rmsd bonds	rmsd angles
50.0–1.75 Å	21.4% / 23.6%	24.3 Å^2	0.005 Å	1.3°

The crystals were in space group I4 with cell dimensions $a = b = 114.13$ Å, $c = 32.59$ Å. *Statistics from SCALA output, CCP4 suite. †The first number refers to overall completeness, the number in brackets refers to the completeness of anomalous data. ‡The number in brackets refers to the highest resolution shell (1.83–1.75 Å). §Statistics from CNS output. #5% of the data were excluded from refinement and were used to determine the R_{free}. The R_{cryst} does not include these reflections. In both cases, $R = \Sigma(|F_{obs}|-k|F_{calc}|)/\Sigma|F_{obs}|$, with an appropriate choice of reflections for the summation.

aIF-5A crystallizes in space group I4 with eight molecules in the unit cell. The crystals are approximately 60% solvent, with wide channels throughout the crystal (Figure 2). aIF-5A is composed almost entirely of β strands and has two domains (Figure 3). The structure shows that the region surrounding the Lys42 is extended from the core of the molecule and that the Lys42 is fully solvent exposed and ready for modification. Joe and Park [13] reported that the core of the N-terminal domain was required for the hypusine modification, that is, recognition of IF-5A by deoxyhypusine synthase, and that small peptides including all of the invariant residues could not be modified. This correlates well with the structure seen, as a small peptide would not be expected to present the Lys42 residue in the same way, as illustrated in Figure 3. This presentation of a lysine residue at the end of a β turn is reminiscent of the biocytin-modified Lys122 of biotin carboxyl carrier protein [14], although the loop of IF-5A is more extended from the core of the protein and there is little sequence similarity between these proteins.

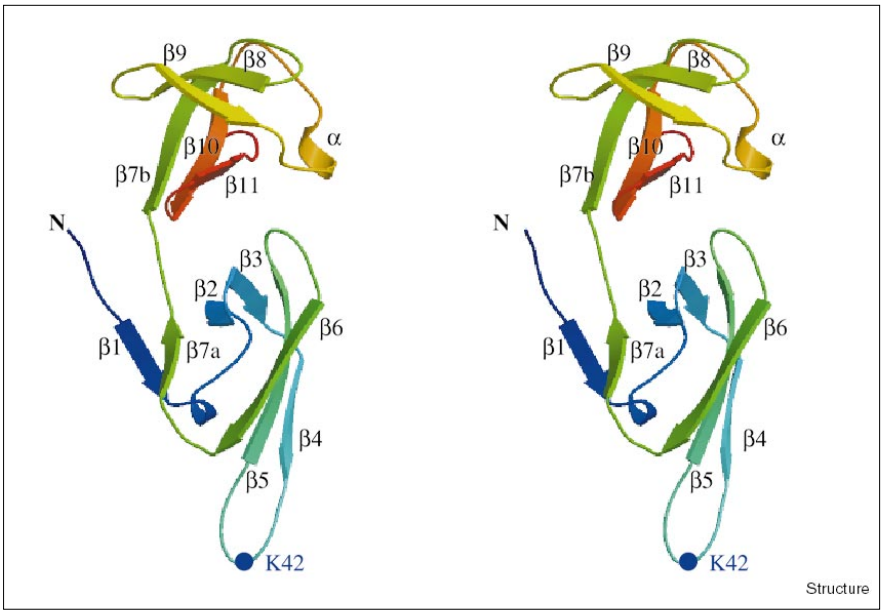
The N-terminal domain contains seven β strands, with β1 and β7a (the N-terminal part of a long strand β7 that extends into the C-terminal domain) hydrogen bonded, the two short strands β2 and β3 hydrogen bonded and β4, β5, β6 forming a sheet on the opposite side of the molecule to β1 and β7a. The β strands are all antiparallel, with β2 and β3 being almost perpendicular to the sheets formed by β1–β7a and β4–β5–β6. There is a short turn of 3_{10}-helix between β1 and β2, residues 13–17. The very long strand β7 extends up to the C-terminal domain and β7b forms a twisted sheet with β10 and β11. This is almost perpendicular to the short strands β8 and β9 on the other side of the domain. The C-terminal domain forms a barrel structure between the sheets and loops. In the loop between β9 and β10 there is a single turn of α helix, containing residues 110–114, (sequence EEEAK). Each domain has a well-defined hydrophobic core, and the contacts between the two domains are extensive and primarily hydrophobic. Phe56 and Val55 of the N-terminal domain are 4–4.5 Å away from Ile128 of the C-terminal domain, and Phe56 is only 3.6 Å from the ring of Tyr108. Tyr22 is 3.7 Å from Trp126, and is less than 4.0 Å from Ile76 in the C-terminal domain. There are also hydrogen bonds between the two domains, for example the Nε of Lys133 is ~2.5 Å from the backbone carbonyl O of Val55.

Figure 2



A space-filling model of 12 aIF-5A molecules in the crystal. The unit cell is defined by the black line box (dimensions $a = b = 114.13$ Å, $c = 32.59$ Å). As can be seen, there are two different channels that run throughout the crystal, one with a diameter of about 30 Å and the other with a 40 Å diameter. Each aIF-5A molecule is colored from the N terminus (dark blue) to the C terminus (red). As can be seen, there are tail–tail interactions as well as head–tail interactions in the crystal. (The figure was generated using the program RasMol [35].)

Figure 3



A stereo diagram of the aIF-5A structure. The molecule is colored from the N terminus (dark blue) to the C terminus (red). The N terminus is denoted in the figure, along with the secondary structure elements (β1 through β11, one 3_{10} helix in the N terminus and one α-helix in the C terminus). The C terminus is hidden behind β strand β10 in this orientation. A blue ball in the loop between strands β4 and β5 denotes the location of the normally hypusinated lysine residue Lys42. (This figure was created using the programs MOLSCRIPT and Raster3D [36,37].)

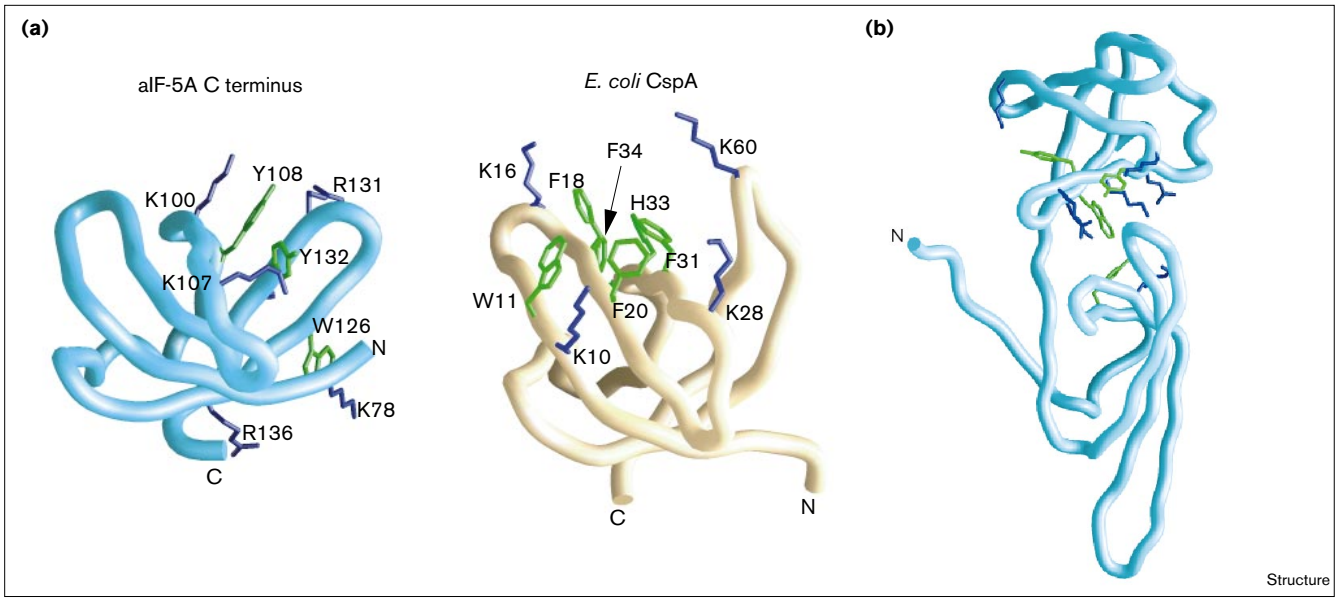
Although there is little sequence similarity between IF-5A and other proteins, each domain of aIF-5A shares some structural homology to previously determined structures. The N-terminal domain (β2 through β7a) is similar in topology to residues 23–76 of dihydrofolate reductase (Brookhaven Protein Data Bank [PDB] code 1vie). 48 Cα atoms superpose with a root mean square deviation (rmsd) of 2.1 Å, although dihydrofolate reductase has a more twisted β structure than IF-5A in this region. The C-terminal domain resembles domains in many proteins, but is most similar to the cold-shock protein CspA from *E. coli* (PDB code 1mjc). The superposition of the Cα atoms of the cold-shock protein and the C-terminal domain (residues 77–138) gives a rmsd of 2.0 Å for 59 residues. The sequence identity between these proteins and aIF-5A is 10% and 12% respectively. Both alignments were found using the DALI program [15]. The cold-shock protein is one of the few proteins in the PDB that has a recognized RNA-binding fold, which is another indication that IF-5A may have a role in RNA metabolism.

The fold of the C-terminal domain and the cold-shock protein is classified as an oligomer-binding (OB) fold which is common among many sugar- and nucleotide-binding proteins [16]. The oligomer-binding site for this fold is often found on one side of the protein, at the end of β3, including the loop between β3 and the α helix, the loop between β1 and β2, and the loop between β4 and β5. This is the area where there are several aromatic and positively charged residues in aIF-5A (Tyr108, Tyr132, Lys100, Lys107, Lys133, and Arg131). There is a second region in aIF-5A that also presents aromatic and positively charged residues on the surface of the protein and this is found in the interface region (Tyr22, Trp126, Lys78, Arg31, and Arg136). These regions differ from those of the known RNA-binding motifs found in the cold-shock protein, which are on the face of the β sheet formed by strands β1–β2–β3 (Figure 4) [17]. Further studies will be needed to tell us which of the two regions on IF-5A is more important for RNA binding.

The N-terminal domain is positively charged and the C-terminal domain is negatively charged (Figure 5). This charge polarity has been noted previously from inspection of the IF-5A sequence alone and may be important for function [13]. IF-5A also has a distinct charge polarization from one face of the molecule to the other, as seen in Figure 5. Residues 4–9 of the N terminus extend from the body of the protein, but the significance of this extension is unclear, as many of the N-terminal residues of human eIF-5A can be dispensed with and the protein can still be hypusinated [13]. We would predict from the structure and sequence alignments that residues 1–8 of the *P. aerophilum* IF-5A protein could be deleted without serious consequences to the modification reaction or to the function of the protein.

Modification by deoxyhypusine synthase requires spermidine as the second substrate [18]. Transfer of the aminobutyl group from spermidine by deoxyhypusine synthase to the specific lysine forms deoxyhypusine, which is subsequently converted by hydroxylation to hypusine by a second specific enzyme, deoxyhypusine hydroxylase. The result of hypusination at position 42 is to extend the lysine residue, effectively doubling its length. As mentioned above, the crystallized protein is not modified. Lys42 is in

Figure 4



The basic and aromatic residues that may be important for RNA binding. **(a)** A comparison of the C-terminal domain of aIF-5A with the cold-shock protein CspA from *E. coli*. The basic and aromatic amino acids thought to be important for RNA binding are shown on a worm representation of the Cα backbone. For the aIF-5A C-terminal domain, residues shown are: Lys78, Lys100, Lys107, Lys133 (not labeled), Arg131, Arg136, Tyr108, Tyr132 and Trp126. For the CspA protein, the residues shown are: Lys10, Lys16, Lys28, Lys60, Trp11, Phe18, Phe20, Phe31, Phe34 and His33. Aromatic residues are in green and basic residues are in blue. **(b)** A worm representation of full-length aIF-5A. The basic and aromatic residues thought to be important for RNA binding form a girdle around the interface between the two domains. The same amino acids are shown as in (a), as well as Arg31 and Tyr22 from the N-terminal domain. The N terminus is denoted for orientation purposes. (These figures were generated using the program GRASP.)

a crystal contact region and addition of the amino-butyl group to Nε of this lysine might very well disrupt the crystal contacts in this region of the crystal. The unmodified protein is found either as a dimer or as a tetramer in solution (as determined by dynamic light scattering, data not shown), but the interactions found within the crystal are almost certainly too tenuous to be relevant to the form found in solution (that is, the buried surface area between molecules in the crystal is only 200 Å^2 to 300 Å^2 per protomer as determined by GRASP [19]).

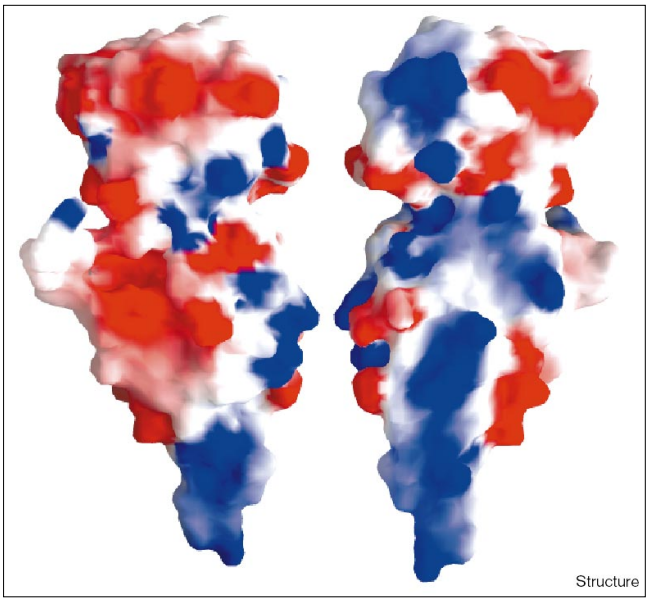
Discussion

The structure of human deoxyhypusine synthase was recently reported by the Davies laboratory [20]. A similar protein is found in the *P. aerophilum* genome; from this and other reports, we would expect that the IF-5A homolog described here would normally be hypusinated *in vivo* in the *P. aerophilum* cell. As discussed previously, hypusination of eukaryotic, and presumably archebacterial, IF-5A is required for function *in vivo* and *in vitro*. Although we would not expect this post-translational modification to alter the structure in any major way, it obviously does alter the functionality of the protein. We would also expect the structure of the *P. aerophilum* deoxyhypusine synthase and the deoxyhypusine hydroxylase proteins to be similar to their counterparts in eukaryotes.

Several specific inhibitors have been found for both deoxyhypusine synthase and deoxyhypusine hydroxylase [21,22]. These inhibitors have been used to show that modification of eIF-5A is necessary for eIF-5A function and have given clues as to that function. Inhibition of modification leads to cell-cycle arrest prior to S-phase entry. IF-5A is linked to the translation of various specific mRNAs involved in cell-cycle regulation and to the regulation of the expression of certain viral structural proteins. How IF-5A regulates the translation of these messages is unknown, but recent experiments have shown that it seems to work downstream of the decapping of mRNAs [9]. Thus, we would expect IF-5A to bind to specific mRNAs and interact with the rRNA or proteins associated with the ribosome.

Strong support for the binding of RNA is also found from the structure itself. The C-terminal domain is distinctly similar to that of known RNA-binding proteins, although there is little sequence homology between these proteins. The RNA-binding motifs RNP1 and RNP2 in the *E. coli* cold-shock protein CspA are found in a valley on one side of the protein [23]. This valley is mainly composed of aromatic residues, is surrounded by several lysines and is a positive region in a distinctly acidic protein (pI ~5.0). A different cleft is found in the C-terminal domain of IF-5A

Figure 5



A surface representation of aIF-5A with the electrostatic potentials shown (± 7 kT). Two views of the molecule are given, with the difference between them being a 180° rotation around the *y* axis (vertical axis). Positively charged residues are shown in blue and negatively charged residues in red. From the figure it is clear that the N-terminal domain is positively charged and that the C-terminal domain is negatively charged, although the overall pI of the protein is calculated to be 6.83 and it has an equal number of acidic and basic amino acids. The figure shows that one side of aIF-5A is mostly negatively charged (left-hand view) whereas the opposite side has a streak of almost continuous positive charge (right-hand view). There is also a girdle around the interface of the two domains that is mostly positively charged. The loop between strands β4 and β5, around the normally hypusinated Lys42 residue, is also very positively charged (bottom of both molecules). The molecule in the left-hand view is in a similar orientation to that of the molecule depicted in Figure 4b. (This figure was generated using the program GRASP.)

and has tyrosine residues surrounded by lysines and an arginine residue. Again, this is a distinctly positive region surrounded by negative charges, as the pI of the C-terminal domain is about 5.3.

IF-5A from eukaryotes has been more extensively studied than the homologs from either archebacteria or eubacteria (IF-P from *E. coli* [24]). The IF-5A homolog in *E. coli* (IF-P) is not hypusinated and is only 20–26% identical to the eukaryotic and archebacterial IF-5A proteins [3,24]. IF-5A proteins are ubiquitous and are found in the cytoplasm of all organisms studied to date. Inhibitors of hypusination block cell growth, although these inhibitors seem to have only a modest effect on general protein synthesis [21]. As few experiments of these kind have been done in archebacteria, we can not say definitively whether the IF-5A homolog has exactly the same role in these organisms.

Biological implications

Translation initiation factor 5A (IF-5A) clearly has a very important role in the life cycle of the cell. Deletion of the gene(s) encoding this protein has been shown to be lethal in all cell types reported to date, including *Escherichia coli*. IF-5A undergoes a post-translational modification by deoxyhypusine synthase that is required for function and cell-cycle progression in eukaryotes. This initiation factor is highly homologous at the amino-acid level in all eukaryotes and this homology extends to the rather diverse archebacterial kingdom. IF-5A is composed of two β-sheet domains, each having structural homology to other proteins in the PDB, despite exhibiting virtually no sequence homology to these other proteins. The structure shows that the IF-5A molecule is charge polarized and that the normally modified lysine is extended away from the main body of the molecule for easy access by deoxyhypusine synthase. Recent reports suggest that IF-5A associates with both mRNA and other factors linked to translation initiation or elongation. IF-5A is also required for the translation of specific mRNAs involved in cell-cycle regulation and specific viral mRNAs from the HIV and T-cell leukemia viruses. The structure of the C-terminal domain is homologous to that of the cold-shock proteins which are known to bind RNA. We would therefore expect that the C-terminal domain is responsible for RNA (either rRNA or mRNA) binding and that the N-terminal domain may be involved in the binding or association of factors involved in translation. The function of the post-translationally modified lysine residue is still unclear, but it must be needed to interact with factors associated with the ribosome.

Materials and methods

Cloning, expression and purification

5′ and 3′ primers were synthesized with *Nde*I and *Bam*HI restriction sites, respectively. These were used with genomic DNA from *Pyrobaculum aerophilum* [25] (gift of Sorel Fitz-Gibbons and Jeffery Miller, UCLA) to amplify out by PCR the full-length ORF containing the coding sequence of gene *pya_828* (IF-5A). The N-terminal leucine residue of aIF-5A was changed to a methionine residue for expression in *E. coli*. This amplicon was enzymatically digested with *Nde*I and *Bam*HI and purified before ligating into a pET28 (Novagen) derivative vector (the pET28 vector had been modified to contain a *Bam*HI site and a 6-His tag at the C terminus). BL21 cells were transformed and clones were screened for expression of the heterologous protein by SDS–PAGE gels. Expression was induced at 37°C with 1 mM IPTG when the cells had reached mid-log phase, and growth was continued for approximately 3.5 h after induction. Protein expression was stopped by centrifugation of the cells and subsequent freezing of the cell pellet at –80°C.

The protein was purified by a simple three-step procedure. The cell pellet from 1 liter of cells was lysed by sonication in approximately 35 ml of low-salt buffer (50 mM Tris pH 8.5, 100 mM NaCl) for 5–10 min in short bursts. The cell debris was removed by centrifugation at 16,500 rpm in a Sorval superspeed centrifuge. The supernatant was removed and heated to ~75°C for 10 min and allowed to cool to about 45°C before centrifuging again to remove the denatured *E. coli* proteins. The supernatant was again removed and this time mixed with

10 ml of Talon bead slurry (Clontech). This slurry was washed five times in high-salt buffer (50 mM Tris pH 8.5, 350 mM NaCl) to remove contaminants and the pure protein was removed from the beads by washing them three times in 10 ml elution buffer (50 mM Tris pH 7.5, 100 mM imidizole, 100 mM NaCl). The elutants were collected and the protein concentrated in Millipore 10 kD filters. The buffer was changed to 20 mM HEPES pH 7.5, 50 mM KCl, 1 mM dithiothreitol (DTT) and the protein concentration was determined spectroscopically at 280 nm to be about 16.5 mg/ml.

Crystallization

Crystallization screens (Hampton) were set up at two temperatures and small crystals were found in different conditions at each temperature. Crystals could be reproducibly grown in 50 mM Hepes pH 7.5, 5–10% PEG 4000 at 8°C when seeded. Crystals grew to approximately 400 x 400 x 50 μm in these conditions with several crystals per drop. Selenomethionine-substituted protein was produced in essentially the same manner except that we used a defined media with added selenomethionine instead of LB when growing up the cells [26]. The substituted protein was purified in the same manner and crystallized in the same conditions, except that 5 mM DTT or 5 mM β-mercaptoethanol was added to keep the selenium reduced. The crystals were cryoprotected with stabilizing solution (50 mM HEPES pH 7.5, 8–12% PEG 4000) fortified with 10% glycerol and 10% ethylene glycol. aIF-5A from *Methanococcus jannaschii* and IF-P from *E. coli* have also been crystallized, although in a different space groups and under different conditions [27,28].

Data collection and structure determination

A three-wavelength data set was collected on a single frozen crystal at beamline x8c at the NSLS. The data were initially reduced and scaled using DENZO and SCALEPACK [29] at beamline x8c, but were later reprocessed at home using MOSFLM and the CCP4 suite [30]. These data were input to the program SOLVE (T.C.T., personal communication) which found and refined three well ordered selenium sites. These three sites were used to phase the reflections and to produce initial maps with an initial figure of merit of 0.65 using data to 2.17 Å. These maps were improved by solvent flattening and histogram matching in DM [30]. The subsequent maps were excellent and the entire polypeptide chain was built from residue 10 to residue 139 using the program O [31]. Residues 4–9 were later added to the model using 2Fo–Fc, Fo–Fc and omit maps.

Refinement

The structure was refined using the impressive CNS program [32]. One round of simulated annealing and nine rounds of minimization with individual temperature factor refinement and occupancy refinement alternated with manual rebuilding resulted in a R_{cryst} of 21.4% and a R_{free} of 23.6% with good stereochemistry. Five percent of the data were reserved for the R_{free} calculation and there was no sigma cutoff made to the data. The data were 99.7% complete in the resolution range of 50.0–1.75 Å and 99.9% complete in the last shell (1.75–1.83 Å). Residue Lys138 is the only residue found outside the allowed Ramachandran plot regions (data from PROCHECK [33]); this is not surprising as it is normally the last amino acid in the protein but has been fused to eight additional amino acids to create the histidine tag used for purification.

Accession numbers

The coordinates and structure factors have been deposited in the PDB under accession numbers 1bkb and r1bkbsf.

Acknowledgements

We thank Michael Altherr, Duilio Casio, Craig Ogata, Roger Sayle and Anthony Nicholls for advice and scientific discussions. We also thank Sorel Fitz-Gibbons and Jeffery Miller for *P. aerophilum* DNA and access to the annotated DNA sequence, and Leon Flaks for assistance at beamline x8c. This work has been supported in part by the Integrated Structural Biology Resource program at LANL.

References

1. Park, M.H., Wolff, E.C. & Folk, J.E. (1993). Is hypusine essential for eukaryotic cell proliferation? *Trends Biochem. Sci.* **18**, 475-479.
2. Magdolen, V., *et al.* & Lottspeich, F. (1994). The function of the hypusine-containing proteins of yeast and other eukaryotes is well conserved. *Mol. Gen. Genet.* **244**, 646-652.
3. Kyrpides, N.C. & Woese, C.R. (1998). Universally conserved translation initiation factors. *Proc. Natl Acad. Sci. USA* **95**, 224-228.
4. Shi, X.P., Yin, K.C., Zimolo, Z.A., Stern, A.M. & Waxman, L. (1996). The subcellular distribution of eukaryotic translation initiation factor eIF-5A in cultured cells. *Exp. Cell Res.* **225**, 348-356.
5. Kemper, W.M., Berry, K.W. & Merrick, W.C. (1976). Purification and properties of rabbit reticulocyte protein synthesis initiation factors M2Bα and M2Bβ. *J. Biol. Chem.* **251**, 5551-5557.
6. Benne, R., Brown-Luedi, M.L. & Hershey, J.W. (1978). Purification and characterization of protein synthesis initiation factors eIF-1, eIF-4C, eIF-4D, and eIF-5 from rabbit reticulocytes. *J. Biol. Chem.* **253**, 3070-3077.
7. Cooper, H.L., Park, M.H. & Folk, J.E. (1982). Posttranslational formation of hypusine in a single major protein occurs generally in growing cells and is associated with activation of lymphocyte growth. *Cell* **29**, 791-797.
8. Schnier, J., Schwelberger, H.G., Smitmcbride, Z., Kang, H.A. & Hershey, J.W.B. (1991). Translation initiation factor 5A and its hypusine modification are essential for cell viability in the yeast *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **11**, 3105-3114.
9. Zuk, D. & Jacobson, A. (1998). A single amino acid substitution in yeast eIF-5A results in messenger RNA stabilization. *EMBO J.* **17**, 2914-2925.
10. Jacobson, A. & Peltz, S.W. (1996). Interrelationships of the pathways of mRNA decay and translation in eukaryotic cells. *Annu. Rev. Biochem.* **65**, 693-740.
11. Hanauske-Abel, H.M., *et al.* & Szabo, P. (1995). Detection of a subset of polysomal messenger RNAs associated with modulation of hypusine formation at the G1-S boundary : proposal of a role for eIF-5A in onset of DNA replication. *FEBS Lett.* **366**, 92-98.
12. Katahira, J., Ishizaki, T., Sakai, H., Adachi, A., Yamamoto, K. & Shida, H. (1995). Effects of translation initiation factor eIF-5A on the functioning of human T-cell leukemia virus type I Rex and human immunodeficiency virus Rev inhibited *trans* dominantly by a Rex mutant deficient in RNA binding. *J. Virol.* **69**, 3125-3133.
13. Joe, Y.A. & Park, M.H. (1994). Structural features of the eIF-5A precursor required for posttranslational synthesis of deoxyhypusine. *J. Biol. Chem.* **269**, 25916-25921.
14. Athappilly, F. & Hendrickson, W. (1995). Structure of the biotinyl domain of acetyl-coenzyme A carboxylase determined by MAD phasing. *Structure* **3**, 1407-1419.
15. Holm, L. & Sander, C. (1993). Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123-138.
16. Murzin, A.G. (1993). OB(oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences. *EMBO J.* **12**, 861-867.
17. Jiang, W., Hou, Y. & Inouye, M. (1997). CspA, the major cold-shock protein of *Escherichia coli* is an RNA chaperone. *J. Biol. Chem.* **272**, 196-202.
18. Park, M.H., Cooper, H.L. & Folk, J.E. (1981). Identification of hypusine: an unusual amino-acid in a protein from human lymphocytes and of spermidine as its biosynthetic precursor. *Proc. Natl Acad. Sci. USA* **78**, 2869-2873.
19. Nicholls, A., Bharadwaj, R. & Honig, B. (1993). GRASP : Graphical Representation and Analysis of Surface-Properties. *Biophysical J.* **64**, A166-A166.
20. Liao, D.I., Wolff, E.C., Park, M.H. & Davies, D.R. (1998). Crystal structure of the NAD complex of human deoxyhypusine synthase : an enzyme with a ball-and-chain mechanism for blocking the active site. *Structure* **6**, 23-32.
21. Park, M.H., Wolff, E.C., Lee, Y.B. & Folk, J.E. (1994). Antiproliferative effects of inhibitors of deoxyhypusine synthase : inhibition of growth of Chinese hamster ovary cells by guanyl diamines. *J. Biol. Chem.* **269**, 27827-27832.
22. Hanauske-Abel, H.M., Park, M.H., Hanauske, A.R., Popowicz, A.M., Lalande, M. & Folk, J.E. (1994). Inhibition of the G1-S transition of the cell cycle by inhibitors of deoxyhypusine hydroxylation. *Biochim. Biophys. Acta* **1221**, 115-124.
23. Schindelin, H., Jiang, W.N., Inouye, M. & Heinemann, U. (1994). Crystal structure of CspA; the major cold shock protein of *Escherichia coli*. *Proc. Nat. Acad. Sci. USA* **91**, 5119-5123.

24. Aoki, H., Dekany, K., Adams, S.L. & Ganoza, M.C. (1997). The gene encoding the elongation factor-P protein is essential for viability and is required for protein synthesis. *J. Biol. Chem.* **272**, 32254-32259.
25. Fitz-Gibbon, S.*, et al.* & Kim, U.-J. (1997). A fosmid-based genomic map and identification of 474 genes of the hyperthermophilic archaeon *Pyrobaculum aerophilum*. *Extremophiles* **1**, 36-51.
26. Doublié, S. (1997). Preparation of selenomethionyl proteins for phase determination. *Methods Enzymol*. Academic Press, Vol. 276, pp. 523-530.
27. Kim, K.K., Yokota, H., Kim, R. & Kim, S.H. (1997) Cloning, expression, and crystallization of a hyperthermophilic protein that is homologous to the eukaryotic translation initiation factor eIF-5A. *Prot. Sci.* **6**, 2268-2270.
28. Aoki, H., Adams, S.L., Turner, M.A. & Ganoza, M.C. (1997). Molecular characterization of the prokaryotic EF-P gene product involved in a peptidyltransferase reaction. *Biochimie* **79**, 7-11.
29. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol*. **276**, 307-326.
30. Bailey, S. (1994). The CCP4 Suite : programs for protein crystallography. *Acta Cryst. D* **50**, 760-763.
31. Jones, T.A., Zou, J.Y., Cowan, S.W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Cryst. A* **47**, 110-119.
32. Brunger, A.T.*, et al.* & Warren, G.L. (1998). Crystallography and NMR system: A new software system for macromolecular structure determination. *Acta Cryst. D*, in press.
33. Laskowski, R.A., Macarthur, M.W., Moss, D.S. & Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283-291.
34. Thompson, J.D., Higgins, D.G. & Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673-4680.
35. Sayle, R. & Milner-White, E.J. (1995). RasMol: Biomolecular graphics for all. *Trends Biochem. Sci.* **20**, 374.
36. Kraulis, P.J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* **24**, 946-950.
37. Merritt, E.A. & Bacon, D.J. (1997). Raster3d : Photorealistic molecular graphics. *Methods Enzymol.* **277**, 505-524.